



## Prediction of Second-Order Moments of Inter-Channel Interference with Principal Component Analysis and Neural Networks

Jones, Rasmus Thomas; Medeiros Diniz, Júlio César; Yankov, Metodi Plamenov; Piels, Molly; Doberstein, Andy; Zibar, Darko

*Published in:*

Proceedings of the 43rd European Conference and Exhibition on Optical Communications (ECOC 2017)

*Link to article, DOI:*

[10.1109/ECOC.2017.8346176](https://doi.org/10.1109/ECOC.2017.8346176)

*Publication date:*

2017

*Document Version*

Peer reviewed version

[Link back to DTU Orbit](#)

*Citation (APA):*

Jones, R. T., Medeiros Diniz, J. C., Yankov, M. P., Piels, M., Doberstein, A., & Zibar, D. (2017). Prediction of Second-Order Moments of Inter-Channel Interference with Principal Component Analysis and Neural Networks. In *Proceedings of the 43rd European Conference and Exhibition on Optical Communications (ECOC 2017)* IEEE. <https://doi.org/10.1109/ECOC.2017.8346176>

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Prediction of Second-Order Moments of Inter-Channel Interference with Principal Component Analysis and Neural Networks

Rasmus T. Jones<sup>(1)</sup>, Júlio C. M. Diniz<sup>(1)</sup>, Metodi P. Yankov<sup>(1)</sup>, Molly Piels<sup>(1)</sup>,  
Andy Doberstein<sup>(2)</sup>, Darko Zibar<sup>(1)</sup>

<sup>(1)</sup> DTU Fotonik, Department Of Photonics Engineering, Technical University of Denmark, Ørstedss Plads, B. 343, 2800 Kgs. Lyngby, Denmark, rajo@fotonik.dtu.dk

<sup>(2)</sup> Keysight Technologies, Harburger Schlossstrasse 6-12, 21079 Hamburg, Germany

**Abstract** A machine learning framework for predicting auto-correlation functions of inter-channel nonlinearities within the uncompensated optical fiber link is proposed. Low generalization error is obtained on the test data.

## Introduction

In wavelength division multiplexed (WDM) optical communication systems intra- and inter-channel nonlinear distortions pose the major limiting factor for increasing the spectral efficiency<sup>1</sup>. Intra-channel nonlinear distortion can be effectively compensated by employing digital back-propagation (DBP). However, inter-channel nonlinear distortions such as cross-phase and cross-polarization modulation (XPM and XPolM) and four wave mixing (FWM) cannot be compensated by DBP due to the absence of the interfering channels (ICs). An approach for compensating and also investigating the impact of inter-channel nonlinearities is to model the nonlinear interference as a stochastic process. An important step in modeling a stochastic process, is identifying its underlying auto-correlation function (ACF). For inter-channel nonlinearities, the ACF is a function of the physical layer parameters such as: Input power, channel spacing, span length and propagation distance.

Inter-channel nonlinearities lead to a long-correlated memory effect. The Gaussian noise (GN) model<sup>2</sup> ignores this effect while proposed models based on perturbation theory describe it<sup>3,4</sup>. The ACF can thereby be computed by employing one of those models. However, the computational time of those models may be prohibitively large if we want to investigate the entire space and all possible combination of the physical layer parameters. A more effective approach is presented in this paper as follows: first a perturbation model is employed for a sparse set of input configurations of the physical layer to emulate the signal propagation through the nonlinear channel, then a machine learning framework<sup>5</sup> is employed to learn the ACF from the data and finally, the learned model is used to predict the ACF for a new set of configuration of the input physical layer parameters. More specifically, we

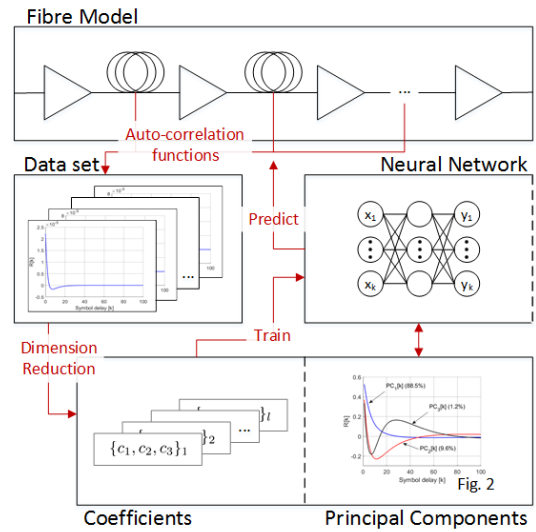
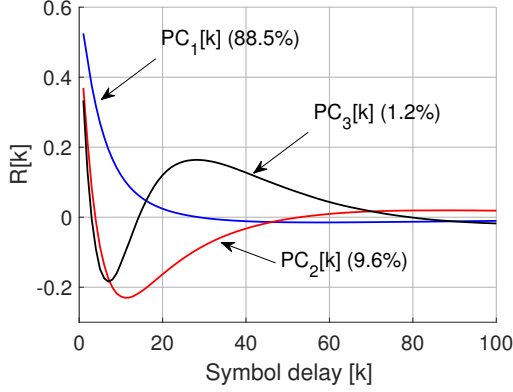


Fig. 1: Setup.

use tools from the machine learning community, such as principal component analysis (PCA) and neural network (NN), to analyze and predict the properties of this long correlated memory effect in terms of ACFs.

## Machine Learning Framework

The employed machine learning framework is shown in Fig. 1 and is described as follows. First, the perturbation model presented by Tao et al.<sup>4</sup> is extended to accommodate for signal propagation through the uncompensated optical fiber link. The data from the perturbation model is then used to compute ACFs. In principle, since the input physical layer parameters are known, one could use supervised learning, such as neural networks, to learn the nonlinear mapping between the input parameters and the generated ACFs. The problem is that the dimensionality (length) of the ACF spans among many symbols. If the entire ACF is to be taken as output (target) values of the neural network, this would result in a highly complex neural network with a large number of trainable



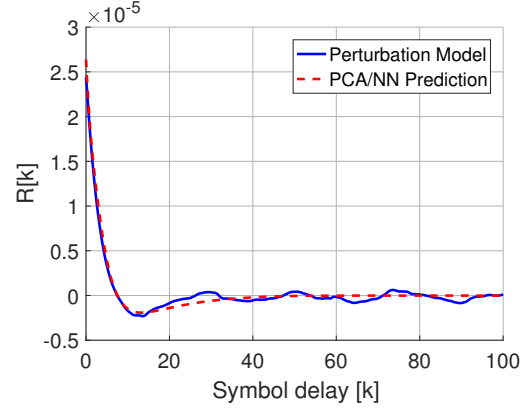
**Fig. 2:** Three principal components ( $PC_d[k]$ ) whose linear combination recover all ACFs with an accuracy of 99.3%. An accuracy of 88.5% is obtained by solely using the first PC.

weights. Instead, it would be favorable to reduce the number of output units of the neural network by transforming the ACFs such that less variables carry most of the information. This is achieved with the PCA, which performs a change of basis on the ACFs. The new basis vectors are called principal components (PCs). They are chosen such that few of them cover most of the variance of the ACFs. With this trick a linear combination of a significant subset of the PCs still approximates the ACFs. Although discarding insignificant PCs introduces information loss, it often reduces the dimensionality of the problem notably. Recovering the ACFs with a subset of  $D = 3$  PCs maintains 99.3% of the variance of the ACFs and is given by

$$R_n[k] = \sum_{d=1}^D c_{n,d} PC_d[k], \quad (1)$$

where  $R_n[k]$  is the  $n$ -th ACF of the data set,  $PC_d[k]$  is the  $d$ -th PC and  $c_{n,d}$  are the coefficients of the PCA. Each ACF  $R_n[k]$  is now described by three coefficients  $\{c_{n,1}, c_{n,2}, c_{n,3}\}$ . The PCs, which represent pseudo ACFs, are shown in Fig. 2. Next, physical layer parameters such as: launch power, channel spacing, span length and propagated distance, are used as an input to the NN and the coefficients of the PCA are used as the output. Thus, the above prohibitive problem is resolved, since the number of NN output units is cut down to 3. Fig. 3, shows one example of a true ACF obtained by the perturbation model and its counterpart obtained by the trained machine learning framework.

Even though we make sure that the most important PCs are used, the discarded PCs also hold information about the ACFs. The information lost during the PCA cannot be recovered by the NN. Meaning, there are two independent sources of error, the dimensionality reduction itself and the possibility of an insufficient NN fit.



**Fig. 3:** ACF as a function of symbol delay, obtained from the model and the machine learning framework. The latter is a linear combination of the three PC in Fig. 2.

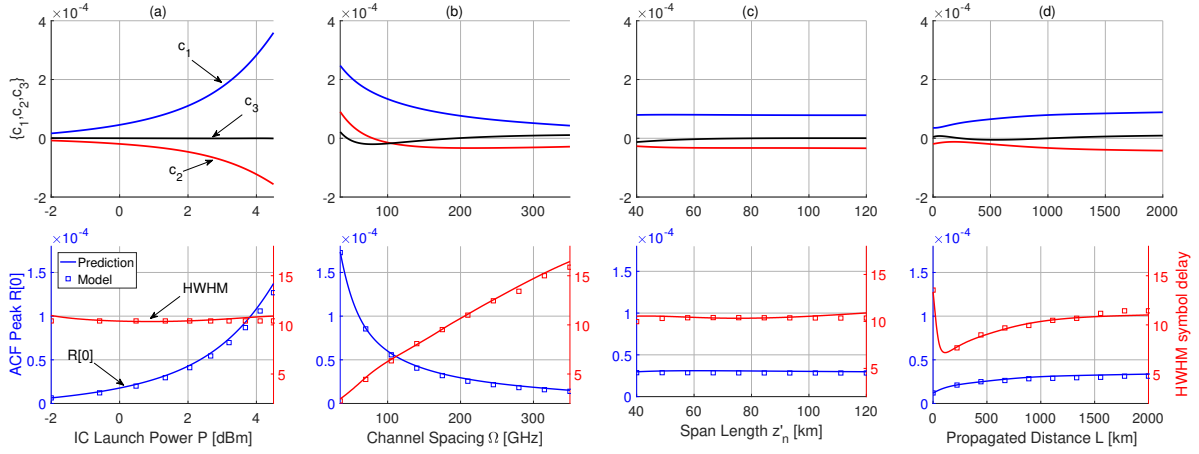
Parameter	Min.	Max.
IC launch power $P$	-2 dBm	4.5 dBm
IC channel spacing $\Omega$	35 GHz	350 GHz
Span length $z'$	40 km	120 km
Propagated distance $L$	0 km	2000 km

**Tab. 1:** Range of physical layer parameters for the generated ACFs

### Training and Prediction

In order to perform the PCA, we generated two data sets of 3000 ACFs, for the XPM and the XPolM processes. In this paper we only show results and figures regarding the XPM process. Differently shaped ACFs are obtained by changing the four physical layer parameters within the ranges given in Tab. 1. Where one ACF describes a XPM process, generated by an IC within a  $z'$  long span, starting at  $L$  into the link, with  $\Omega$  channel spacing and a launch power of  $P$ . Other parameters, such as the modulation format, sample frequency, CD coefficient  $D$ , nonlinear coefficient  $\gamma$ , attenuation  $\alpha$  and roll-off factor are set to 16 QAM, 32 GHz, 17 ps/(km nm),  $1.2 \text{ (W km)}^{-1}$ , 0.2 dB/km and 0.1, respectively.

The NN is trained to learn a continuous nonlinear mapping from the input space of physical layer parameters  $\{P, \Omega, z', L\}$  to the output space of three coefficients  $\{c_1, c_2, c_3\}$  determining the actual ACF. During the training of a NN the normalised data set is randomly split up in training and test data. While the training data is used to learn the nonlinear mapping, the test data is used to evaluate the robustness of the fit. The NN produces a good fit when the mean squared error (MSE) is low. Especially the MSE of the test set, since these ACFs have not been seen by the NN during training. A MSE of  $6.4 \cdot 10^{-3}$  and  $7.8 \cdot 10^{-3}$  are reported on the training and test set, respectively, which shows the high accuracy which can be achieved with the NN. It also justifies using the NN for the continuous analysis below.



**Fig. 4:** (top) PC coefficients  $\{c_1, c_2, c_3\}$  measuring the contribution of their respective PC to the overall ACF as a function of the physical layer parameters. (bottom) The ACF peak  $R[0]$  and the HWHM as a function of the physical layer parameters for model (squares) and prediction (solid lines).

## Analysis

The PCA extracts the main features of the ACFs from the data set as PCs, see Fig. 2. The first PC  $PC_1[k]$  is most important. It depicts a quickly decaying always positive ACF with short tail. This feature models the power of the XPM process. The second and third PCs,  $PC_2[k]$  and  $PC_3[k]$ , model the long tail caused by CD. Since the PCs are multiplied by the coefficients, the sign of the coefficients are important. For instance, a positive coefficient  $c_2$  and its PC results in a correlation function with local minima, whereas a negative  $c_2$  results in a local maxima. Further, the reconstruction of arbitrary ACFs also enables the analysis of the peak at  $R[0]$  and the half width half maximum (HWHM) of the ACF.

Every column in Fig. 4 shows how the contribution of the PC coefficients change (top) and how the peak  $R[0]$  and HWHM of the ACF changes (bottom) with continuously changing the parameters. When sweeping one of the four parameters the others are set to the center of their respective range.

Column (a) considers the IC launch power. More power results in increased absolute value of the contribution of all coefficients in all cases. The first PC is effected more drastically which is associated with its feature of modeling the power. This is also reflected in the increasing peak  $R[0]$  of the ACF. Column (b) depicts the dependency of the channel spacing. Here, ICs which are further away from the channel of interest interfere less. The coefficients  $c_2$  and  $c_3$  change their sign with increasing channel spacing at around 100 GHz. The HWHM value shows that this behaviour of the coefficients contributes to a longer tail in the resulting ACF. This is explained by the longer walk-off of the IC with higher channel spacing. Column (c) shows that the span length has minor effects on the ACF, especially because all nonlinear interactions occur in the beginning of the span

where the power is still high. Increasing the span length beyond a certain point has no effect as the IC does not interact anymore. Column (d) shows that without CD there is hardly any memory and with increasing CD the three coefficients and the HWHM find themselves on saturated levels.

## Conclusion

We have proposed a method of determining complex high dimensional properties of the nonlinear fiber channel, where an analytical description is not available. Based on a training set, the PCA enables us to extract the most important features of the ACF of the nonlinear interference and the NN predicts the solution in between data points. Together they become a tool to analyze the effects on a system when physical layer parameters are changing. Further research includes how sparse the data set may be while maintaining robust predictions at the same time.

## Acknowledgements

This work was financially supported by Keysight Technologies (Germany, Böblingen).

## References

- [1] R.-J. Essiambre et al., "Capacity Limits of Optical Fiber Networks." JLT, **28**(4) p.662 (2010)
- [2] P. Poggiolini et. al., "Recent Advances in the Modeling of the Impact of Non-Linear Fiber Propagation Effects on Un-compensated Coherent Transmission Systems." JLT, **35**(3) p.458 (2017)
- [3] R. Dar et al., "Inter-channel Nonlinear Interference Noise in WDM Systems: Modeling and Mitigation." JLT, **33**(5) p.1044 (2015)
- [4] Z. Tao et al., "Simple Fiber Model for Determination of XPM Effects." JLT, **29**(7) p.974 (2011)
- [5] C. Bishop, "Pattern Recognition and Machine Learning (Information Science and Statistics)." Springer, (2006)